Attorney's Docket No.: 07039-454001

APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: ARSENIC METHYLTRANSFERASE SEQUENCE VARIANTS

APPLICANTS: BIANCA A. THOMAE, ERIC D. WIEBEN, RICHARD M. WEINSHILBOUM, AND THOMAS C. WOOD

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EV 321179991 US

April 15, 2004

Date of Deposit

ARSENIC METHYLTRANSFERASE SEQUENCE VARIANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from U.S. Provisional Application Serial No. 60/463,114, filed April 15, 2003.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

Funding for the work described herein was provided in part by the federal government under grant nos. RO1 GM28157, RO1 GM35720, and UO1 GM61388. The federal government may have certain rights in the invention.

TECHNICAL FIELD

The invention relates to arsenic methyltransferase (ASMT) nucleic acid and amino acid sequence variants.

BACKGROUND

Acute exposure to inorganic arsenic compounds can lead to fever, cardiac arrhythmia, cardiac failure, hepatomegaly, melanosis, peripheral neuropathy, hematopoietic effects, loss of peripheral nervous system function, leukopenia, anemia, or death. Chronic exposure can lead to neurotoxicity, demyelination, liver injury, peripheral vascular disease, and carcinogenesis resulting in hemangiosarcoma of liver, skin cancer, and lung cancer. The majority of occupational exposure to arsenic is in the manufacture of pesticides, herbicides, and other agricultural products, and in the smelting industry. Exposure to arsenic also can result from environmental exposure to contaminated ground water. Metabolism of arsenic is complex as arsenic can be trivalent or pentavalent and can form many different compounds. Methylated and dimethylated arsenic compounds are the major transformation products *in vivo* and are rapidly excreted in urine. While methylation typically is regarded as a mechanism for detoxification, certain methylated arsenic compounds that contain As^{III} are more cytotoxic and genotoxic than arsenate (the most stable form of arsenic) and arsenite (AsO_3^{3-}), and also more potent inhibitors of GSH reductase, thioredoxin reductase, and pyruvate dehydrogenase than arsenite. See, Lin et al., J. Biol. Chem. 277(13):10795-10803 (2002). ASMT (also referred to as AMT)

is an enzyme that methylates arsenite using S-adenosyl-L-methionine as the methyl group donor. ASMT is expressed in the liver, kidney, and brain in humans. In rats, ASMT is expressed in heart, adrenal glands, urinary bladder, brain, kidney, lung, and liver.

SUMMARY

The invention is based on the discovery of sequence variants that occur in both coding and non-coding regions of *ASMT* nucleic acids. Certain *ASMT* nucleotide sequence variants encode ASMT enzymes that are associated with individual differences in enzymatic activity. Other sequence variants in non-coding regions of the *ASMT* nucleic acid may alter regulation of transcription and/or splicing of the *ASMT* nucleic acid. Discovery of these sequence variants allows individual differences in the methylation of drugs and other xenobiotics in humans to be assessed such that particular treatment regimens can be tailored to an individual based on the presence or absence of one or more sequence variants. Identification of *ASMT* sequence variants also allows predisposition to hemangiosarcoma of liver, skin cancer, and lung cancer to be assessed in individuals.

In one aspect, the invention features an isolated nucleic acid molecule containing an *ASMT* nucleic acid sequence, wherein the nucleic acid molecule is at least ten nucleotides in length, and wherein the *ASMT* nucleic acid sequence comprises a nucleotide sequence variant. The nucleotide sequence variant can be at a position selected from the group consisting of position 2278, 2412, 2477, 2534, 2615, 2838, 2840, 3370, 3398, 3435, 5791, 6176, 6324, 6373, 6426, 8011, 8078, 10259, 12025, 12084, 12327, 23855, 23936, 33672, 33765, and 33860 of SEQ ID NO:1.

The nucleotide sequence variant can be a nucleotide substitution or a nucleotide insertion. For example, the nucleotide sequence variant can be a cytosine substitution for thymine at position 2278 of SEQ ID NO:1; an adenine substitution for guanine at position 2412 of SEQ ID NO:1; a guanine substitution for adenine at position 2477 of SEQ ID NO:1; a guanine substitution for cytosine at position 2534 of SEQ ID NO:1; a cytosine substitution for thymine at position 2615 of SEQ ID NO:1; an adenine substitution for cytosine at position 2838 of SEQ ID NO:1; or a cytosine substitution for guanine at position 2840 of SEQ ID NO:1. The nucleotide sequence variant also can be an adenine substitution for thymine at nucleotide 3370 of position 3370 of SEQ ID NO:1; an

insertion of a cytosine at position 3398 of SEQ ID NO:1; or a thymine substitution for guanine at position 3435 of SEQ ID NO:1. The nucleotide sequence variant can be an adenine substitution for guanine at position 5791 of SEQ ID NO:1; a guanine substitution for an adenine at position 6178 of SEQ ID NO:1; an adenine substitution for a guanine at position 6324 of SEQ ID NO:1; a cytosine substitution for thymine at position 6373 of SEQ ID NO:1; or a thymine substitution for adenine at position 6426 of SEQ ID NO:1. The nucleotide sequence variant can be a thymine substitution for cytosine at position 8011 of SEQ ID NO:1; a guanine substitution for adenine at position 8078 of SEQ ID NO:1; a cytosine substitution for guanine at position 10259 of SEQ ID NO:1; a cytosine substitution for an adenine at position 12025 of SEQ ID NO:1; or a thymine substitution for a cytosine at position 12084 of SEQ ID NO:1. The nucleotide sequence variant can be a cytosine substitution for thymine at position 12327 of SEQ ID NO:1, a cytosine substitution for thymine at position 23855 of SEQ ID NO:1; or a thymine substitution for cytosine at position 23936 of SEQ ID NO:1. The nucleotide sequence variant also can be a thymine substitution for cytosine at position 33672 of SEQ ID NO:1, an adenine substitution for guanine at position 33765 of SEQ ID NO:1, or an adenine substitution for guanine at position 33860 of SEQ ID NO:1.

Alternatively, the variant can be an insertion or a deletion of a variable number tandem repeat. The deletion or insertion can be between nucleotides 2820 and 3020 of SEQ ID NO:1.

In another aspect, the invention features an isolated nucleic acid encoding an ASMT polypeptide, wherein the polypeptide contains an ASMT amino acid sequence variant relative to the amino acid sequence of SEQ ID NO:5. The amino acid sequence variant can be at a residue selected from the group consisting of 173, 287, and 306 (e.g., a tryptophan at residue 173, a threonine at residue 287, or an isoleucine at residue 306).

In another aspect, the invention features an isolated ASMT polypeptide, wherein the polypeptide contains an ASMT amino acid sequence variant relative to the amino acid sequence of SEQ ID NO:5. The amino acid sequence variant can be at a residue selected from the group consisting of 173, 287, and 306 (e.g., a tryptophan at residue 173, a threonine at residue 287, or an isoleucine at residue 306). Activity of the polypeptide can be altered relative to a wild type ASMT polypeptide.

The invention also features an isolated nucleic acid molecule containing an *ASMT* nucleic acid sequence, wherein the nucleic acid molecule is at least ten nucleotides in length, wherein the *ASMT* nucleic acid sequence has at least 99% sequence identity to a region of SEQ ID NO:3. In the *ASMT* nucleic acid sequence, position 594 is a thymine, position 937 is a cytosine, and position 994 is a thymine. The region can be selected from the group consisting of nucleotides 550 to 650 of SEQ ID NO:3, nucleotides 900 to 950 of SEQ ID NO:3, and nucleotides 951 to 1000 of SEQ ID NO:3.

In yet another aspect, the invention features an article of manufacture including a substrate, wherein the substrate includes a population of isolated *ASMT* nucleic acid molecules, and wherein the nucleic acid molecules include an *ASMT* nucleotide sequence variant. The substrate can include a plurality of discrete regions, wherein each region includes a different population of isolated *ASMT* nucleic acid molecules, and wherein each population of molecules includes a different *ASMT* nucleotide sequence variant.

The invention also features a method for determining if a mammal is predisposed to increased risk for acute or chronic arsenic exposure. The method includes obtaining a biological sample from a mammal, and detecting the presence or absence of an *ASMT* nucleotide sequence variant in the sample, wherein risk for toxicity is determined based on the presence or absence of a variant. The method can further include detecting the presence or absence of a plurality of *ASMT* nucleotide sequence variants in the sample to obtain a variant profile of the mammal, and wherein risk for toxicity is determined based on the variant profile.

In another aspect, the invention features a method for assisting a medical or research professional. The method includes obtaining a biological sample from a mammal, and detecting the presence or absence of a plurality of *ASMT* nucleotide sequence variants in the sample to obtain a variant profile of the mammal. The method can further include communicating the profile to the medical or research professional.

In yet another aspect, the invention features a method for determining the methyltransferase status of an individual. The method includes determining whether the subject contains a variant *ASMT* nucleic acid.

The invention also features an isolated nucleic acid molecule including an *ASMT* nucleic acid sequence, wherein the nucleic acid molecule is at least ten nucleotides in

length, and wherein the *ASMT* nucleic acid sequence includes at least two nucleotide sequence variants. The variants can be within any combination of coding sequences, intron sequences, 5' untranslated sequences, or 3' untranslated sequences.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DESCRIPTION OF DRAWINGS

Figure 1 is the nucleotide sequence of the reference *ASMT* (SEQ ID NO:1) and its complement (SEQ ID NO:2). Exons are labeled and are in bold type. Single nucleotide polymorphisms (SNPs) are circled and labeled. Primer sequences are underlined, and the start and stop codons are double-underlined. The translation initiation codon begins at nucleotide 2954 of SEQ ID NO:1. Exon 1 contains nucleotides 2877 to 2954 of SEQ ID NO:1. Intron 1 contains nucleotides 2955 to 3165 of SEQ ID NO:1. Exon 2 contains nucleotides 3166 to 3206 of SEQ ID NO:1. Intron 2 contains nucleotides 3207 to 3444 of SEQ ID NO:1. Exon 3 contains nucleotides 3445 to 3572 of SEQ ID NO:1. Intron 3 contains nucleotides 3573 to 5808 of SEQ ID NO:1. Exon 4 contains nucleotides 5809 to 5959 of SEQ ID NO:1. Intron 4 contains nucleotides 5960 to 6457 of SEQ ID NO:1. Exon 5 contains nucleotides 6458 to 6594 of SEQ ID NO:1. Intron 5 contains nucleotides 6595 to 7952 of SEQ ID NO:1. Exon 6 contains nucleotides 7953 to 8022 of SEQ ID NO:1. Intron 6 contains nucleotides 8023 to 10314 of SEQ ID NO:1. Exon 7 contains nucleotides 10315 to 10396 of SEQ ID NO:1. Intron 7 contains nucleotides 10397 to 11739 of SEQ ID NO:1. Exon 8 contains nucleotides 11740 to 11871 of SEQ ID NO:1. Intron 8 contains nucleotides 11872 to 12209 of SEQ ID NO:1. Exon 9

contains nucleotides 12210 to 12352 of SEQ ID NO:1. Intron 9 contains nucleotides 12353 to 23904 of SEQ ID NO:1. Exon 10 contains nucleotides 23905 to 24039 of SEQ ID NO:1. Intron 10 contains nucleotides 24040 to 33953 of SEQ ID NO:1. Exon 11 contains nucleotides 33954 to 34161 of SEQ ID NO:1.

Figure 2A is a cDNA sequence (SEQ ID NO:3) containing the open reading frame of the reference *ASMT* (nucleotides 78-1202) and the complementary sequence (SEQ ID NO:4) of the cDNA sequence. SNPs are circled, and the start and stop codons are double-underlined. Figure 2B is the amino acid sequence (SEQ ID NO:5) of the reference ASMT.

Figure 3 is a schematic of the locations of polymorphisms within the human *ASMT* sequence in Caucasian Americans (CA) and African Americans (AA).

Figure 4A is a graph showing levels of luciferase activity in extracts from HEK293 cells transfected with the indicated *ASMT* reporter plasmids. Figure 4B is a graph showing levels of luciferase activity in extracts from HepG2 cells transfected with the indicated *ASMT* reporter plasmids.

DETAILED DESCRIPTION

The invention features *ASMT* nucleotide and amino acid sequence variants. ASMT is an enzyme that methylates arsenite and methylarsonous acid using S-adenosylmethionine (SAM) as the methyl donor. Genetically-based variations in ASMT that lead to altered levels of ASMT or altered ASMT activity may be important in determining the risk associated with acute or chronic exposure to arsenic and development of arsenic-induced skin lesions, cancer (e.g., liver, skin, or lung), neurotoxicity, neuropathy, or liver injury.

Nucleic Acid Molecules

The invention features isolated nucleic acids that include an *ASMT* nucleic acid sequence. The *ASMT* nucleic acid sequence includes a nucleotide sequence variant and nucleotides flanking the sequence variant. As used herein, "isolated nucleic acid" refers to a nucleic acid that is separated from other nucleic acid molecules that are present in a mammalian genome, including nucleic acids that normally flank one or both sides of the nucleic acid in a mammalian genome (e.g., nucleic acids that encode non-ASMT

proteins). The term "isolated" as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, lentivirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a recombinant DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not to be considered an isolated nucleic acid.

Nucleic acids of the invention are at least about 8 nucleotides in length. For example, the nucleic acid can be about 8, 9, 10-20 (e.g., 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length), 20-50, 50-100 or greater than 100 nucleotides in length (e.g., greater than 150, 200, 250, 300, 350, 400, 450, 500, 750, or 1000 nucleotides in length). Nucleic acids of the invention can be in a sense or antisense orientation, can be complementary to the *ASMT* reference sequence (e.g., SEQ ID NO:2 and SEQ ID NO:4), and can be DNA, RNA, or nucleic acid analogs. Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxyuridine for deoxythymidine, and 5-methyl-2'-deoxycytidine or 5-bromo-2'-doxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six-membered, morpholino ring, or peptide

nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. See, Summerton and Weller, Antisense Nucleic Acid Drug Dev. (1997) 7(3):187-195; and Hyrup et al. (1996) Bioorgan. Med. Chem. 4(1):5-23. In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

As used herein, "nucleotide sequence variant" refers to any alteration in an *ASMT* reference sequence, and includes variations that occur in coding and non-coding regions, including exons, introns, and untranslated sequences. Nucleotides are referred to herein by the standard one-letter designation (A, C, G, or T). Variations include single nucleotide substitutions, deletions of one or more nucleotides, and insertions of one or more nucleotides. The reference *ASMT* nucleic acid sequence is provided in Figure 1 (SEQ ID NO:1) and in GenBank (Accession No. NT_008804). The reference *ASMT* cDNA including the *ASMT* ORF is provided in Figure 2A (SEQ ID NO:3) and the corresponding reference ASMT amino acid sequence is provided in Figure 2B (SEQ ID NO:5). The nucleic acid and amino acid reference sequences also are referred to herein as "wild type."

As used herein, "untranslated sequence" includes 5' and 3' flanking regions that are outside of the messenger RNA (mRNA) as well as 5' and 3' untranslated regions (5'-UTR or 3'-UTR) that are part of the mRNA, but are not translated. Positions of nucleotide sequence variants in 5' untranslated sequences are designated as "-X" relative to the "A" in the translation initiation codon; positions of nucleotide sequence variants in the coding sequence and 3' untranslated sequence are designated as "+X" or "X" relative to the "A" in the translation initiation codon. Nucleotide sequence variants that occur in introns are designated as "+X" or "X" relative to the "G" in the splice donor site (GT) or as "-X" relative to the "G" in the splice acceptor site (AG).

In some embodiments, an *ASMT* nucleotide sequence variant encodes an ASMT polypeptide having an altered amino acid sequence. The term "polypeptide" refers to a chain of at least four amino acid residues (e.g., 4-8, 9-12, 13-15, 16-18, 19-21, 22-100, 100-150, 150-200, 200-250 residues, or a full-length ASMT polypeptide). ASMT polypeptides may or may not have ASMT catalytic activity, or may have altered activity

relative to the reference ASMT polypeptide. Polypeptides that do not have activity or have altered activity are useful for diagnostic purposes (e.g., for producing antibodies having specific binding affinity for variant ASMT polypeptides).

Corresponding ASMT polypeptides, irrespective of length, that differ in amino acid sequence are herein referred to as allozymes. For example, an *ASMT* nucleic acid sequence that includes a thymine at nucleotide 517 (nucleotide 8011 of SEQ ID NO:1) encodes an ASMT polypeptide having a tryptophan at amino acid residue 173. This polypeptide (Arg173Trp) would be considered an allozyme with respect to the reference ASMT polypeptide that contains an arginine at amino acid residue 173. Additional non-limiting examples of *ASMT* sequence variants that alter amino acid sequence include variants at nucleotides 860 and 917 (nucleotides 12327 and 23936, respectively, of SEQ ID NO:1). For example, an *ASMT* nucleic acid molecule can include a cytosine at nucleotide 860 and encode an ASMT polypeptide having a threonine at amino acid residue 287 in place of a methionine residue (Met287Thr); or a thymine at nucleotide 917 and encode an ASMT polypeptide having an isoleucine at amino acid 306 in place of a threonine residue (Thr306Ile).

ASMT allozymes as described herein are encoded by a series of ASMT alleles. These alleles represent nucleic acid sequences containing sequence variants, typically multiple sequence variants, within coding and non-coding sequences. Representative examples of single nucleotide variants are described above. Table 2 sets out a series of *ASMT* alleles that encode ASMT. Some alleles are commonly observed, i.e., have allele frequencies >1%, such as the allele having a guanine at nucleotide -477 (nucleotide 2477 of SEQ ID NO:1) in place of an adenine. The relatively large number of alleles and allozymes for ASMT indicates the potential complexity of ASMT pharmacogenetics. Such complexity emphasizes the need for determining single nucleotide variants, (i.e., single nucleotide polymorphisms, SNPs) as well as multiple nucleotide variants and complete *ASMT* haplotypes (i.e., the set of alleles on one chromosome or a part of a chromosome) of patients. See, e.g., the haplotypes set forth in Table 5.

Certain *ASMT* nucleotide sequence variants do not alter the amino acid sequence. Such variants, however, could alter regulation of transcription as well as mRNA stability.

ASMT variants can occur in intron sequences, for example, within introns 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. See, for example, the intronic sequence variants set forth in Table 2.

ASMT nucleotide sequence variants that do not change the amino acid sequence also can be within an exon or in 5' or 3' untranslated sequences. Nucleotide sequence variants in the 5' UTR can include a cytosine substitution for thymine at nucleotide -676 (nucleotide 2278 of SEQ ID NO:1); an adenine substitution for guanine at nucleotide -542 (nucleotide 2412 of SEQ ID NO:1); a guanine substitution for adenine at nucleotide -477 (nucleotide 2477 of SEQ ID NO:1); a cytosine substitution for thymine at nucleotide -339 (nucleotide 2615 of SEQ ID NO:1); an adenine substitution for cytosine at nucleotide -116 (nucleotide 2838 of SEQ ID NO:1); or a cytosine substitution for guanine at nucleotide -114 (nucleotide 2840 of SEQ ID NO:1).

Other variants in the 5' untranslated sequences can be an insertion or deletion of one or more variable number tandem repeats (VNTR). A VNTR can be any tandemly repeated sequence. Typically, a VNTR can be between about 20 and about 50 (e.g., 20, 22, 24, 26, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 44, 46, 48, or 50) nucleotides in length. For example, a VNTR can contain between about 30 and about 40 (e.g., 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40) contiguous nucleotides of the sequence 5'-GAGTCGCAGGCCGAGGAGACAGTGAGTGCGCGCCCTGAGT-3' (SEQ ID NO:6). Alternatively, a VNTR can have a sequence that between about 30 and about 40 nucleotides in length and is at least 90% identical (e.g., 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical) to the sequence set forth in SEQ ID NO:6, wherein the percent identity is determined as described below. A VNTR can be located in the 5' flanking region, exon 1, intron 1, and/or combinations thereof. Thus, the insertion or deletion of a VNTR can be in the 5' untranslated region between, for example, nucleotides 2820 and 3020 of SEQ ID NO:1 (e.g., between nucleotides 2830 and 3010, 2840 and 3000, or 2850 and 2990 of SEQ ID NO:1). In one embodiment, a VNTR can have a nucleotide sequence that is 36 nucleotides in length and contains *ASMT* sequences from the 5'-FR and exon 1. In another embodiment, a VNTR can have a nucleotide sequence that is 35 nucleotides in length and contains *ASMT* sequences from exon 1 and intron 1. A genomic *ASMT* nucleic acid sequence typically can include two, three, or four VNTRs. While a change in the number of VNTRs does not alter the

encoded ASMT amino acid sequence, an increase or a decrease in the number of repeats can increase or decrease expression of an ASMT polypeptide.

In some embodiments, nucleic acid molecules of the invention can have at least 97% (e.g., 97.5%, 98%, 98.5%, 99.0%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity with a region of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4 that includes one or more variants described herein. The region of SEQ ID NO:1, 2, 3, or 4 is at least ten nucleotides in length (e.g., 10, 15, 20, 50, 60, 70, 75, 100, 150 or more nucleotides in length). For example, a nucleic acid molecule can have at least 99% identity with nucleotides 550 to 650, 900 to 950, 925 to 975, 951 to 1000, or 950 to 1050 of SEQ ID NO:3, where the nucleotide sequence of SEQ ID NO:3 includes one or more of the variants described herein. For example, the nucleotide sequence of SEQ ID NO:3 can have a thymine at nucleotide 594, a cytosine at nucleotide 937, or a thymine at nucleotide 994, and combinations thereof.

Percent sequence identity is calculated by determining the number of matched positions in aligned nucleic acid sequences, dividing the number of matched positions by the total number of aligned nucleotides, and multiplying by 100. A matched position refers to a position in which identical nucleotides occur at the same position in aligned nucleic acid sequences. Percent sequence identity also can be determined for any amino acid sequence. To determine percent sequence identity, a target nucleic acid or amino acid sequence is compared to the identified nucleic acid or amino acid sequence using the BLAST 2 Sequences (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTN version 2.0.14 and BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (World Wide Web at "fr" dot "com" slash "blast") or the U.S. government's National Center for Biotechnology Information web site (World Wide Web at "ncbi" dot "nlm" dot "nih" dot "gov"). Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ.

Bl2seq performs a comparison between two sequences using either the BLASTN or BLASTP algorithm. BLASTN is used to compare nucleic acid sequences, while BLASTP is used to compare amino acid sequences. To compare two nucleic acid sequences, the options are set as follows: -i is set to a file containing the first nucleic acid

sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastn; -o is set to any desired file name (e.g., C:\output.txt); -q is set to -1; -r is set to 2; and all other options are left at their default setting. The following command will generate an output file containing a comparison between two sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2. If the target sequence shares homology with any portion of the identified sequence, then the designated output file will present those regions of homology as aligned sequences. If the target sequence does not share homology with any portion of the identified sequence, then the designated output file will not present aligned sequences.

Once aligned, a length is determined by counting the number of consecutive nucleotides from the target sequence presented in alignment with sequence from the identified sequence starting with any matched position and ending with any other matched position. A matched position is any position where an identical nucleotide is presented in both the target and identified sequence. Gaps presented in the target sequence are not counted since gaps are not nucleotides. Likewise, gaps presented in the identified sequence are not counted since target sequence nucleotides are counted, not nucleotides from the identified sequence.

The percent identity over a particular length is determined by counting the number of matched positions over that length and dividing that number by the length followed by multiplying the resulting value by 100. For example, if (1) a 1000 nucleotide target sequence is compared to the sequence set forth in SEQ ID NO:1, (2) the Bl2seq program presents 969 nucleotides from the target sequence aligned with a region of the sequence set forth in SEQ ID NO: 1 where the first and last nucleotides of that 969 nucleotide region are matches, and (3) the number of matches over those 969 aligned nucleotides is 900, then the 1000 nucleotide target sequence contains a length of 969 and a percent identity over that length of 93 (i.e., 900 ÷ 969 x 100 = 93).

It will be appreciated that different regions within a single nucleic acid target sequence that aligns with an identified sequence can each have their own percent identity. It is noted that the percent identity value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 are rounded down to 78.1, while 78.15, 78.16, 78.17,

78.18, and 78.19 are rounded up to 78.2. It also is noted that the length value will always be an integer.

Isolated nucleic acid molecules of the invention can be produced by standard techniques, including, without limitation, common molecular cloning and chemical nucleic acid synthesis techniques. For example, polymerase chain reaction (PCR) techniques can be used to obtain an isolated nucleic acid containing an *ASMT* nucleotide sequence variant. PCR refers to a procedure or technique in which target nucleic acids are enzymatically amplified. Sequence information from the ends of the region of interest or beyond typically is employed to design oligonucleotide primers that are identical in sequence to opposite strands of the template to be amplified. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA. Primers are typically 14 to 40 nucleotides in length, but can range from 10 nucleotides to hundreds of nucleotides in length. General PCR techniques are described, for example in PCR Primer: A Laboratory Manual, ed. by Dieffenbach and Dveksler, Cold Spring Harbor Laboratory Press, 1995. When using RNA as a source of template, reverse transcriptase can be used to synthesize complementary DNA (cDNA) strands. Ligase chain reaction, strand displacement amplification, self-sustained sequence replication, or nucleic acid sequence-based amplification also can be used to obtain isolated nucleic acids. See, for example, Lewis Genetic Engineering News, 12(9):1 (1992); Guatelli et al., Proc. Natl. Acad. Sci. USA, 87:1874-1878 (1990); and Weiss, Science, 254:1292 (1991).

Isolated nucleic acids of the invention also can be chemically synthesized, either as a single nucleic acid molecule (e.g., using automated DNA synthesis in the 3' to 5' direction using phosphoramidite technology) or as a series of oligonucleotides. For example, one or more pairs of long oligonucleotides (e.g., >100 nucleotides) can be synthesized that contain the desired sequence, with each pair containing a short segment of complementarity (e.g., about 15 nucleotides) such that a duplex is formed when the oligonucleotide pair is annealed. DNA polymerase is used to extend the oligonucleotides, resulting in a single, double-stranded nucleic acid molecule per oligonucleotide pair, which then can be ligated into a vector.

Isolated nucleic acids of the invention also can be obtained by mutagenesis. For example, the reference sequences depicted in Figures 1 or 2A can be mutated using standard techniques including oligonucleotide-directed mutagenesis and site-directed mutagenesis through PCR. See, Short Protocols in Molecular Biology, Chapter 8, Green Publishing Associates and John Wiley & Sons, edited by Ausubel et al., 1992. Examples of positions that can be modified are described herein.

ASMT Polypeptides

Isolated ASMT polypeptides of the invention include an amino acid sequence variant relative to the reference ASMT (Figure 2B, SEQ ID NO:5). The term "isolated" with respect to an ASMT polypeptide refers to a polypeptide that has been separated from cellular components by which it is naturally accompanied. Typically, the polypeptide is isolated when it is at least 60% (e.g., 70%, 80%, 90%, 95%, or 99%), by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. In general, an isolated polypeptide will yield a single major band on a non-reducing polyacrylamide gel.

ASMT polypeptides of the invention include variants at one or more of amino acid residues 173, 287, and 306. In particular, a tryptophan residue can be substituted at position 173, a threonine residue at position 287, or an isoleucine residue at position 306. In some embodiments, activity of ASMT polypeptides is altered relative to the reference ASMT. Certain ASMT allozymes can have reduced activity, while other allozymes can have activity that is comparable to the reference ASMT. Other allozymes can have increased activity relative to the reference ASMT. Activity of ASMT polypeptides can be assessed *in vitro*. For example, the activity of ASMT polypeptides can be assessed by determining the amount of [^{14}C]-methylated arsenic products that are produced by a recombinant methyltransferase (e.g., recombinant ASMT) in the presence of sodium arsenite (2.5 mM) and ^{14}C-SAM (10 μM).

Other biochemical properties of allozymes, such as apparent K_m values, also can be altered relative to the reference ASMT. Apparent K_m values can be calculated, for example, by using the method of Wilkinson with a computer program written by Cleland. Wilkinson, Biochem. J., 80:324-332 (1961); and Cleland, Nature, 198:463-365 (1963).

Isolated polypeptides of the invention can be obtained, for example, by extraction from a natural source (e.g., brain tissue), chemical synthesis, or by recombinant production in a host cell. To recombinantly produce ASMT polypeptides, a nucleic acid encoding an *ASMT* nucleotide sequence variant can be ligated into an expression vector and used to transform a prokaryotic (e.g., bacteria) or eukaryotic (e.g., insect, yeast, or mammal) host cell. In general, nucleic acid constructs include a regulatory sequence operably linked to an *ASMT* nucleic acid sequence. Regulatory sequences (e.g., promoters, enhancers, polyadenylation signals, or terminators) do not typically encode a gene product, but instead affect the expression of the nucleic acid sequence. In addition, a construct can include a tag sequence designed to facilitate subsequent manipulations of the expressed nucleic acid sequence (e.g., purification, localization). Tag sequences, such as green fluorescent protein (GFP), glutathione S-transferase (GST), six histidine (His_6), c-myc, hemagglutinin, or Flag™ tag (Kodak) sequences are typically expressed as a fusion with the expressed nucleic acid sequence. Such tags can be inserted anywhere within the polypeptide including at either the carboxyl or amino termini. The type and combination of regulatory and tag sequences can vary with each particular host, cloning or expression system, and desired outcome. A variety of cloning and expression vectors containing combinations of regulatory and tag sequences are commercially available. Suitable cloning vectors include, without limitation, pUC18, pUC19, and pBR322 and derivatives thereof (New England Biolabs, Beverly, MA), and pGEN (Promega, Madison, WI). Additionally, representative prokaryotic expression vectors include pBAD (Invitrogen, Carlsbad, CA), the pTYB family of vectors (New England Biolabs), and pGEMEX vectors (Promega); representative mammalian expression vectors include pTet-On/pTet-Off (Clontech, Palo Alto, CA), pIND, pVAX1, pCR3.1, pcDNA3.1, pcDNA4, or pUni (Invitrogen), and pCI or pSI (Promega); representative insect expression vectors include pBacPAK8 or pBacPAK9 (Clontech), and p2Bac (Invitrogen); and representative yeast expression vectors include MATCHMAKER (Clontech) and pPICZ A, B, and C (Invitrogen).

In bacterial systems, a strain of *Escherichia coli* can be used to express ASMT variant polypeptides. For example, BL-21 cells can be transformed with a pGEX vector containing an *ASMT* nucleic acid sequence. The transformed bacteria can be grown

exponentially and then stimulated with isopropylthiogalactopyranoside (IPTG) prior to harvesting. In general, the ASMT-GST fusion proteins produced from the pGEX expression vector are soluble and can be purified easily from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the expressed ASMT polypeptide can be released from the GST moiety.

In eukaryotic host cells, a number of viral-based expression systems can be utilized to express ASMT variant polypeptides. A nucleic acid encoding a polypeptide of the invention can be cloned into, for example, a baculoviral vector such as pBlueBac (Invitrogen) and then used to co-transfect insect cells such as *Spodoptera frugiperda* (Sf9) cells with wild type DNA from *Autographa californica* multinuclear polyhedrosis virus (AcMNPV). Recombinant viruses producing polypeptides of the invention can be identified by standard methodology. Alternatively, a nucleic acid encoding a polypeptide of the invention can be introduced into a SV40, retroviral, or vaccinia based viral vector and used to infect suitable host cells.

Eukaryotic cell lines that stably express ASMT variant polypeptides can be produced using expression vectors with the appropriate control elements and a selectable marker. For example, the eukaryotic expression vector pCR3.1 (Invitrogen, San Diego, CA) and p91023(B) (see Wong et al., Science (1985) 228:810-815) or modified derivatives thereof are suitable for expression of ASMT variant polypeptides in, for example, Chinese hamster ovary (CHO) cells, COS-1 cells, human embryonic kidney 293 cells, NIH3T3 cells, BHK21 cells, MDCK cells, and human vascular endothelial cells (HUVEC). Following introduction of the expression vector by electroporation, lipofection, calcium phosphate or calcium chloride co-precipitation, DEAE dextran, or other suitable transfection method, stable cell lines are selected, e.g., by antibiotic resistance to G418, kanamycin, or hygromycin. Alternatively, amplified sequences can be ligated into a eukaryotic expression vector such as pcDNA3 (Invitrogen) and then transcribed and translated *in vitro* using wheat germ extract or rabbit reticulocyte lysate.

ASMT variant polypeptides can be purified by known chromatographic methods including ion exchange and gel filtration chromatography. See, for example, Caine et al., Protein Expr. Purif. (1996) 8(2):159-166. ASMT polypeptides can be "engineered" to

contain a tag sequence describe herein that allows the polypeptide to be purified (e.g., captured onto an affinity matrix). Immunoaffinity chromatography also can be used to purify ASMT polypeptides.

Non-Human Mammals

The invention features non-human mammals that include *ASMT* nucleic acids of the invention, as well as progeny and cells of such non-human mammals. Non-human mammals include, for example, rodents such as rats, guinea pigs, and mice, and farm animals such as pigs, sheep, goats, horses, and cattle. Non-human mammals of the invention can express an *ASMT* variant nucleic acid in addition to an endogenous *ASMT* (e.g., a transgenic non-human that includes an *ASMT* nucleic acid randomly integrated into the genome of the non-human mammal). Alternatively, an endogenous *ASMT* nucleic acid can be replaced with an *ASMT* variant nucleic acid of the invention by homologous recombination. See, Shastry, Mol. Cell Biochem., (1998) 181(1-2):163-179, for a review of gene targeting technology.

In one embodiment, non-human mammals are produced that lack an endogenous *ASMT* nucleic acid (i.e., a knockout), and then an *ASMT* variant nucleic acid of the invention is introduced into the knockout non-human mammal. Nucleic acid constructs used for producing knockout non-human mammals can include a nucleic acid sequence encoding a selectable marker, which is generally used to interrupt the targeted exon site by homologous recombination. Typically, the selectable marker is flanked by sequences homologous to the sequences flanking the desired insertion site. It is not necessary for the flanking sequences to be immediately adjacent to the desired insertion site. Suitable markers for positive drug selection include, for example, the aminoglycoside 3N phosphotransferase gene that imparts resistance to geneticin (G418, an aminoglycoside antibiotic), and other antibiotic resistance markers, such as the hygromycin-B-phosphotransferase gene that imparts hygromycin resistance. Other selection systems include negative-selection markers such as the thymidine kinase (TK) gene from herpes simplex virus. Constructs utilizing both positive and negative drug selection also can be used. For example, a construct can contain the aminoglycoside phosphotransferase gene

and the TK gene. In this system, cells are selected that are resistant to G418 and sensitive to gancyclovir.

To create non-human mammals having a particular gene inactivated in all cells, it is necessary to introduce a knockout construct into the germ cells (sperm or eggs, i.e., the "germ line") of the desired species. Genes or other DNA sequences can be introduced into the pronuclei of fertilized eggs by microinjection. Following pronuclear fusion, the developing embryo may carry the introduced gene in all its somatic and germ cells because the zygote is the mitotic progenitor of all cells in the embryo. Since targeted insertion of a knockout construct is a relatively rare event, it is desirable to generate and screen a large number of animals when employing such an approach. Because of this, it can be advantageous to work with the large cell populations and selection criteria that are characteristic of cultured cell systems. However, for production of knockout animals from an initial population of cultured cells, it is necessary that a cultured cell containing the desired knockout construct be capable of generating a whole animal. This is generally accomplished by placing the cell into a developing embryo environment of some sort.

Cells capable of giving rise to at least several differentiated cell types are "pluripotent." Pluripotent cells capable of giving rise to all cell types of an embryo, including germ cells, are hereinafter termed "totipotent" cells. Totipotent murine cell lines (embryonic stem, or "ES" cells) have been isolated by culture of cells derived from very young embryos (blastocysts). Such cells are capable, upon incorporation into an embryo, of differentiating into all cell types, including germ cells, and can be employed to generate animals lacking an endogenous *ASMT* nucleic acid. That is, cultured ES cells can be transformed with a knockout construct and cells selected in which the *ASMT* gene is inactivated.

Nucleic acid constructs can be introduced into ES cells, for example, by electroporation or other standard technique. Selected cells can be screened for gene targeting events. For example, the polymerase chain reaction (PCR) can be used to confirm the presence of the transgene.

The ES cells further can be characterized to determine the number of targeting events. For example, genomic DNA can be harvested from ES cells and used for

Southern analysis. See, for example, Section 9.37-9.52 of Sambrook et al., Molecular Cloning, A Laboratory Manual, second edition, Cold Spring Harbor Press, Plainview; NY, 1989.

To generate a knockout animal, ES cells having at least one inactivated *ASMT* allele are incorporated into a developing embryo. This can be accomplished through injection into the blastocyst cavity of a murine blastocyst-stage embryo, by injection into a morula-stage embryo, by co-culture of ES cells with a morula-stage embryo, or through fusion of the ES cell with an enucleated zygote. The resulting embryo is raised to sexual maturity and bred in order to obtain animals, whose cells (including germ cells) carry the inactivated *ASMT* allele. If the original ES cell was heterozygous for the inactivated *ASMT* allele, several of these animals can be bred with each other in order to generate animals homozygous for the inactivated allele.

Alternatively, direct microinjection of DNA into eggs can be used to avoid the manipulations required to turn a cultured cell into an animal. Fertilized eggs are totipotent, i.e., capable of developing into an adult without further substantive manipulation other than implantation into a surrogate mother. To enhance the probability of homologous recombination when eggs are directly injected with knockout constructs, it is useful to incorporate at least about 8 kb of homologous DNA into the targeting construct. In addition, it is also useful to prepare the knockout constructs from isogenic DNA.

Embryos derived from microinjected eggs can be screened for homologous recombination events in several ways. For example, if the *ASMT* gene is interrupted by a coding region that produces a detectable (e.g., fluorescent) gene product, then the injected eggs are cultured to the blastocyst stage and analyzed for presence of the indicator polypeptide. Embryos with fluorescing cells, for example, are then implanted into a surrogate mother and allowed to develop to term. Alternatively, injected eggs are allowed to develop and DNA from the resulting pups analyzed by PCR or RT-PCR for evidence of homologous recombination.

Nuclear transplantation also can be used to generate non-human mammals of the invention. For example, fetal fibroblasts can be genetically modified such that they contain an inactivated endogenous *ASMT* gene and express an *ASMT* nucleic acid of the

invention, and then fused with enucleated oocytes. After activation of the oocytes, the eggs are cultured to the blastocyst stage, and implanted into a recipient. See, Cibelli et al., Science, (1998) 280:1256-1258. Adult somatic cells, including, for example, cumulus cells and mammary cells, can be used to produce animals such as mice and sheep, respectively. See, for example, Wakayama et al., Nature, (1998) 394(6691):369-374; and Wilmut et al., Nature, (1997) 385(6619):810-813. Nuclei can be removed from genetically modified adult somatic cells, and transplanted into enucleated oocytes. After activation, the eggs can be cultured to the 2-8 cell stage, or to the blastocyst stage, and implanted into a suitable recipient. Wakayama et al. 1998, *supra*.

Non-human mammals of the invention such as mice can be used, for example, to screen toxicity of compounds that are substrates for ASMT, drugs that alter ASMT activity, or for carcinogenesis. For example, ASMT activity or toxicity can be assessed in a first group of such non-human mammals in the presence of a compound, and compared with ASMT activity or toxicity in a corresponding control group in the absence of the compound. As used herein, suitable compounds include biological macromolecules such as an oligonucleotide (RNA or DNA), or a polypeptide of any length, a chemical compound, a mixture of chemical compounds, or an extract isolated from bacterial, plant, fungal, or animal matter. The concentration of compound to be tested depends on the type of compound and *in vitro* test data.

Non-human mammals can be exposed to test compounds by any route of administration, including enterally (e.g., orally) and parenterally (e.g., subcutaneously, intravascularly, intramuscularly, or intranasally). Suitable formulations for oral administration can include tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulfate). Tablets can be coated by methods known in the art. Preparations for oral administration can also be formulated to give controlled release of the compound.

Compounds can be prepared for parenteral administration in liquid form (e.g., solutions, solvents, suspensions, and emulsions) including sterile aqueous or non-aqueous carriers. Aqueous carriers include, without limitation, water, alcohol, saline, and buffered solutions. Examples of non-aqueous carriers include, without limitation, propylene glycol, polyethylene glycol, vegetable oils, and injectable organic esters. Preservatives and other additives such as, for example, antimicrobials, anti-oxidants, chelating agents, inert gases, and the like may also be present. Pharmaceutically acceptable carriers for intravenous administration include solutions containing pharmaceutically acceptable salts or sugars. Intranasal preparations can be presented in a liquid form (e.g., nasal drops or aerosols) or as a dry product (e.g., a powder). Both liquid and dry nasal preparations can be administered using a suitable inhalation device. Nebulised aqueous suspensions or solutions can also be prepared with or without a suitable pH and/or tonicity adjustment.

Detecting ASMT Sequence Variants

ASMT nucleotide sequence variants can be detected, for example, by sequencing exons, introns, 5' untranslated sequences, or 3' untranslated sequences, by performing allele-specific hybridization, allele-specific restriction digests, mutation specific polymerase chain reactions (MSPCR), by single-stranded conformational polymorphism (SSCP) detection (Schafer et al., 1995, Nat. Biotechnol. 15:33-39), denaturing high performance liquid chromatography (DHPLC, Underhill et al., 1997, Genome Res., 7:996-1005), infrared matrix-assisted laser desorption/ionization (IR-MALDI) mass spectrometry (WO 99/57318), and combinations of such methods.

Genomic DNA generally is used in the analysis of *ASMT* nucleotide sequence variants, although mRNA also can be used. Genomic DNA is typically extracted from a biological sample such as a peripheral blood sample, but can be extracted from other biological samples, including tissues (e.g., mucosal scrapings of the lining of the mouth or from renal or hepatic tissue). Routine methods can be used to extract genomic DNA from a blood or tissue sample, including, for example, phenol extraction. Alternatively, genomic DNA can be extracted with kits such as the QIAamp® Tissue Kit (Qiagen,

Chatsworth, CA), Wizard® Genomic DNA purification kit (Promega) and the A.S.A.P.™ Genomic DNA isolation kit (Boehringer Mannheim, Indianapolis, IN).

Typically, an amplification step is performed before proceeding with the detection method. For example, exons or introns of the *ASMT* gene can be amplified then directly sequenced. Dye primer sequencing can be used to increase the accuracy of detecting heterozygous samples.

Allele specific hybridization also can be used to detect sequence variants, including complete haplotypes of a mammal. See, Stoneking et al., 1991, Am. J. Hum. Genet. 48:370-382; and Prince et al., 2001, Genome Res., 11(1):152-162. In practice, samples of DNA or RNA from one or more mammals can be amplified using pairs of primers and the resulting amplification products can be immobilized on a substrate (e.g., in discrete regions). Hybridization conditions are selected such that a nucleic acid probe can specifically bind to the sequence of interest, e.g., the variant nucleic acid sequence. Such hybridizations typically are performed under high stringency as some sequence variants include only a single nucleotide difference. High stringency conditions can include the use of low ionic strength solutions and high temperatures for washing. For example, nucleic acid molecules can be hybridized at 42°C in 2X SSC (0.3M NaCl/0.03 M sodium citrate/0.1% sodium dodecyl sulfate (SDS) and washed in 0.1X SSC (0.015M NaCl/0.0015 M sodium citrate), 0.1% SDS at 65°C. Hybridization conditions can be adjusted to account for unique features of the nucleic acid molecule, including length and sequence composition. Probes can be labeled (e.g., fluorescently) to facilitate detection. In some embodiments, one of the primers used in the amplification reaction is biotinylated (e.g., 5' end of reverse primer) and the resulting biotinylated amplification product is immobilized on an avidin or streptavidin coated substrate.

Allele-specific restriction digests can be performed in the following manner. For nucleotide sequence variants that introduce a restriction site, restriction digest with the particular restriction enzyme can differentiate the alleles. For *ASMT* sequence variants that do not alter a common restriction site, mutagenic primers can be designed that introduce a restriction site when the variant allele is present or when the wild type allele is present. A portion of *ASMT* nucleic acid can be amplified using the mutagenic primer and a wild type primer, followed by digest with the appropriate restriction endonuclease.

Certain variants, such as insertions or deletions of one or more nucleotides, change the size of the DNA fragment encompassing the variant. The insertion or deletion of nucleotides can be assessed by amplifying the region encompassing the variant and determining the size of the amplified products in comparison with size standards. For example, a region of *ASMT* can be amplified using a primer set from either side of the variant. One of the primers is typically labeled, for example, with a fluorescent moiety, to facilitate sizing. The amplified products can be electrophoresed through acrylamide gels with a set of size standards that are labeled with a fluorescent moiety that differs from the primer.

PCR conditions and primers can be developed that amplify a product only when the variant allele is present or only when the wild type allele is present (MSPCR or allele-specific PCR). For example, patient DNA and a control can be amplified separately using either a wild type primer or a primer specific for the variant allele. Each set of reactions is then examined for the presence of amplification products using standard methods to visualize the DNA. For example, the reactions can be electrophoresed through an agarose gel and the DNA visualized by staining with ethidium bromide or other DNA intercalating dye. In DNA samples from heterozygous patients, reaction products would be detected with each set of primers. Patient samples containing solely the wild type allele would have amplification products only in the reaction using the wild type primer. Similarly, patient samples containing solely the variant allele would have amplification products only in the reaction using the variant primer. Allele-specific PCR also can be performed using allele-specific primers that introduce priming sites for two universal energy-transfer-labeled primers (e.g., one primer labeled with a green dye such as fluoroscein and one primer labeled with a red dye such as sulforhodamine). Amplification products can be analyzed for green and red fluorescence in a plate reader. See, Myakishev et al., 2001, Genome 11(1):163-169.

Mismatch cleavage methods also can be used to detect differing sequences by PCR amplification, followed by hybridization with the wild type sequence and cleavage at points of mismatch. Chemical reagents, such as carbodiimide or hydroxylamine and osmium tetroxide can be used to modify mismatched nucleotides to facilitate cleavage.

Alternatively, ASMT variants can be detected by antibodies that have specific binding affinity for variant ASMT polypeptides. Variant ASMT polypeptides can be produced in various ways, including recombinantly, as discussed above. Host animals such as rabbits, chickens, mice, guinea pigs, and rats can be immunized by injection of an ASMT variant polypeptide. Various adjuvants that can be used to increase the immunological response depend on the host species and include Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Polyclonal antibodies are heterogeneous populations of antibody molecules that are contained in the sera of the immunized animals. Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be prepared using an ASMT variant polypeptide and standard hybridoma technology. In particular, monoclonal antibodies can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture such as described by Kohler et al., Nature, 256:495 (1975), the human B-cell hybridoma technique (Kosbor et al., Immunology Today, 4:72 (1983); Cole et al., Proc. Natl. Acad. Sci USA, 80:2026 (1983)), and the EBV-hybridoma technique (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 (1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the monoclonal antibodies of the invention can be cultivated *in vitro* and *in vivo*.

Antibody fragments that have specific binding affinity for an ASMT variant polypeptide can be generated by known techniques. For example, such fragments include but are not limited to F(ab')2 fragments that can be produced by pepsin digestion of the antibody molecule, and Fab fragments that can be generated by reducing the disulfide bridges of F(ab')2 fragments. Alternatively, Fab expression libraries can be constructed. See, for example, Huse et al., Science, 246:1275 (1989). Once produced, antibodies or fragments thereof are tested for recognition of ASMT variant polypeptides by standard immunoassay methods including ELISA techniques, radioimmunoassays and Western blotting. See, Short Protocols in Molecular Biology, Chapter 11, Green Publishing Associates and John Wiley & Sons, edited by Ausubel et al., 1992.

Methods

As a result of the present invention, it is possible to determine methyltransferase status of a subject (e.g., a mammal such as a human). "Methyltransferase status" refers to the ability of a subject to transfer a methyl group to a substrate (e.g., arsenite). Methyltransferase status of a subject can be determined by measuring the level of methyltransferase (e.g., ASMT) activity in the subject using, for example, the methods described herein. Alternatively, methyltransferase status can be evaluated by determining whether a methyltransferase nucleic acid sequence (e.g., an *ASMT* nucleic acid sequence) of a subject contains one or more variants (e.g., one or more variants that are correlated with increased or decreased methyltransferase activity). A variant that results in decreased or increased ASMT activity, for example, can be said to result in "reduced" or "enhanced" methyltransferase status, respectively. In some embodiments, the variant profile of a subject can be used to determine the methyltransferase status of the subject.

"Variant profile" refers to the presence or absence of a plurality (e.g., two or more) of *ASMT* nucleotide sequence variants or ASMT amino acid sequence variants. For example, a variant profile can include the complete *ASMT* haplotype of the subject (e.g., see Table 5) or can include the presence or absence of a set of particular non-synonymous SNPs (e.g., single nucleotide substitutions that alter the amino acid sequence of an ASMT polypeptide). In one embodiment, determining the variant profile includes detecting the presence or absence of two or more non-synonymous SNPs (e.g., 2, 3, or 4 non-synonymous SNPs), including those described herein. There may be ethnic-specific pharmacogenetic variation, as certain of the nucleotide and amino acid sequence variants described herein were detected solely in African-American or Caucasian-American subjects. In addition, determining the variant profile can include detecting the presence or absence of any type of *ASMT* variant (e.g., a SNP or an alteration in the number of VNTRs) together with any other *ASMT* variant (e.g., a polymorphism pair or a group of polymorphism pairs). Such polymorphism pairs include, without limitation, the pairs described in Table 4. Further, determining the variant profile can include detecting the presence or absence of any *ASMT* variant together with one or more variants from other methyltransferases.

Methyltransferase activity of an enzyme such as ASMT can be measured using, for example, *in vitro* methods such as those described herein. As used herein, the term "reduced methyltransferase status" refers to a decrease (e.g., a 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, 95%, or 100% decrease) in methyltransferase activity (e.g., ASMT activity) of a subject, as compared to a control level of methyltransferase activity. Similarly, the term "enhanced methyltransferase status" refers to an increase (e.g., a 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, 95%, 100%, or more than 100% increase) in methyltransferase activity of a subject, as compared to a control level of methyltransferase activity. A control level of methyltransferase activity can be, for example, an average level of methyltransferase activity in a population of individuals. In one embodiment, the population includes individuals that do not contain particular *ASMT* nucleotide sequence variants or particular ASMT amino acid sequence variants (e.g., particular variants that affect methyltransferase status). Alternatively, a control level of methyltransferase activity can refer to the level of methyltransferase activity in a control subject (e.g., a subject that does not contain an *ASMT* nucleic acid containing a variant).

In further embodiments of the invention, methyltransferase status can be linked to predisposition to (i.e., a relative greater risk of) a particular condition (e.g., acute or chronic toxicity from arsenic exposure). Additional risk factors including, for example, family history and other genetic factors (e.g., polymorphisms in reductases that convert arsenate and methylarsonic acid to arsine and methylarsonous acid) can be considered when determining risk. Predisposition to such conditions can be determined based on the presence or absence of a single *ASMT* sequence variant or based on a variant profile.

Articles of Manufacture

Articles of manufacture of the invention include populations of isolated *ASMT* nucleic acid molecules or ASMT polypeptides immobilized on a substrate. Suitable substrates provide a base for the immobilization of the nucleic acids or polypeptides, and in some embodiments, allow immobilization of nucleic acids or polypeptides into discrete regions. In embodiments in which the substrate includes a plurality of discrete regions, different populations of isolated nucleic acids or polypeptides can be immobilized in each

discrete region. Thus, each discrete region of the substrate can include a different *ASMT* nucleic acid or ASMT polypeptide sequence variant. Such articles of manufacture can include two or more sequence variants of ASMT, or can include all of the sequence variants known for ASMT. For example, the article of manufacture can include two or more of the sequence variants identified herein and one or more other ASMT sequence variants, such as nucleic acid variants that occur in the 5'-flanking region of the *ASMT* gene. Furthermore, nucleic acid molecules containing sequence variants for other methyltransferases can be included on the substrate.

Suitable substrates can be of any shape or form and can be constructed from, for example, glass, silicon, metal, plastic, cellulose, or a composite. For example, a suitable substrate can include a multiwell plate or membrane, a glass slide, a chip, or polystyrene or magnetic beads. Nucleic acid molecules or polypeptides can be synthesized *in situ,* immobilized directly on the substrate, or immobilized via a linker, including by covalent, ionic, or physical linkage. Linkers for immobilizing nucleic acids and polypeptides, including reversible or cleavable linkers, are known in the art. See, for example, U.S. Patent No. 5,451,683 and WO98/20019. Immobilized nucleic acid molecules are typically about 20 nucleotides in length, but can vary from about 10 nucleotides to about 1000 nucleotides in length.

In practice, a sample of DNA or RNA from a subject can be amplified, the amplification product hybridized to an article of manufacture containing populations of isolated nucleic acid molecules in discrete regions, and hybridization can be detected. Typically, the amplified product is labeled to facilitate detection of hybridization. See, for example, Hacia et al., Nature Genet., 14:441-447 (1996); and U.S. Patent Nos. 5,770,722 and 5,733,729.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1 – Methods and Materials

PCR Amplification and DNA Sequencing: The gene encoding ASMT (also known as CYT19) was cloned based on homology with the rat and mouse sequences and other known methyltransferases. Tissue localization was determined by 3' and 5' RACE.

Anonymized DNA samples from 60 Caucasian-American and 60 African-American subjects were obtained from the Coriell Institute Cell Repository (Camden, NJ). Eleven PCR reactions were performed with each DNA sample to amplify all *ASMT* exons and splice junctions. The amplicons were then sequenced using dye-primer sequencing chemistry to facilitate the identification of heterozygous bases (Chadwick et al. Biotechniques 20:676-683 (1996)). Universal M13 sequencing tags were added to the 5'-ends of each forward and reverse primer for sequencing purposes. All forward primers contained the M13 forward sequence (5'-TGTAAAACGACGGCCAGT-3'; SEQ ID NO:7), and all reverse primers contained the M13 reverse sequence (5'-CAGGAAACAGCTATGACC-3'; SEQ ID NO:8). The sequences and locations of each primer within the gene are listed in Table 1. "F" represents forward; "R", reverse; "U", upstream; "D", downstream; "I", intron; "FR", flanking region; and "UTR", untranslated region. The locations of primers in the gene were chosen to avoid repetitive sequence.

Amplifications were performed with AmpliTaq Gold DNA polymerase (Perkin Elmer, Foster City, CA) using a "hot start" to help ensure amplification specificity. Amplicons were sequenced in the Mayo Molecular Biology Core Facility with an ABI 377 DNA sequencer using BigDye™ (Perkin Elmer) dye-primer sequencing chemistry. Both DNA strands were sequenced in all cases. To exclude PCR-induced artifacts, independent amplification followed by DNA sequencing was performed for all samples in which a SNP was only observed once among the samples resequenced. DNA sequence chromatograms were analyzed using the PolyPhred 3.0 (Nickerson et al. Nucl. Acids Res. 25:2745-2751 (1997)) and Consed 8.0 (Gordon et al. Genome Res. 8:195-202 (1998)) programs developed by the University of Washington (Seattle, WA). The University of Wisconsin GCG software package, Version 10, was also used to analyze nucleotide sequence. GenBank accession numbers for the *ASMT* reference sequences were AF226730.

Table 1

PCR primers used for resequencing ASMT

Primer Name	Primer Location	Primer Sequence (5' to 3' direction)*	SEQ ID NO:
UF(-783) M13	5'-FR	TGTAAAACGACGGCCAGTTCTGCAAGATGAAGTGACAATAC	9
UR(-275) M13	5'-FR	CAGGAAACAGCTATGACCTTGTTCGCTCCACTGCGATT	10
UF(-312) M13	5'-FR	TGTAAAACGACGGCCAGTACGAGATTTATCCGTGAAAATCGCA	11
I1R64 M13	Intron 1	CAGGAAACAGCTATGACCGGAGGCGCTCGCGTTCCCGCCTT	12
I1F86 M13	Intron 1	TGTAAAACGACGGCCAGTAGCTGTGTCTCGAGACCTTTGT	13
I3R84 M13	Intron 3	CAGGAAACAGCTATGACCGTTAATTCCCTAGTGACCTGCATCATTAT	14
I3F(-256)M13	Intron 3	TGTAAAACGACGGCCAGTTTATTTTAGATGGCTTATGAAGTCTTAGT	15
I4R172 M13	Intron 4	CAGGAAACAGCTATGACCTAATTGTTCAAGTTATCAGTTTCCAA	16
I4F101 M13	Intron 4	TGTAAAACGACGGCCAGTAGCTTCTAGTTAGCAATGCTCATTT	17
I5R33 M13	Intron 5	CAGGAAACAGCTATGACCGCTATAGTCATAACAGTAAGAATATAGA	18
I5F(-234)M13	Intron 5	TGTAAAACGACGGCCAGTTGCCTGATGTTATTTTCTGCACATTCAAACTT	19
I6R224 M13	Intron 6	CAGGAAACAGCTATGACCATTCAGTGACAACTGTCACCACGGATTTA	20
I6F(-179)M13	Intron 6	TGTAAAACGACGGCCAGTCCTTGGCACTTGACTATTGATTGTAA	21
I7R78M13	Intron 7	CAGGAAACAGCTATGACCATCCTGGCTATTAGCAGAAAGGAGTT	22
I7F(-277) M13	Intron 7	TGTAAAACGACGGCCAGTGTGGCGTACTGCATACATGAATATTAT	23
I8R115 M13	Intron 8	CAGGAAACAGCTATGACCCCAGATAACAGAAAGTATCCCTCAAT	24
I8F9 M13	Intron 8	TGTAAAACGACGGCCAGTTGACAGACAGCAGGGACTATTATAA	25
I9R96 M13	Intron 9	CAGGAAACAGCTATGACCAGAAAAATGGGAGGCAATGCAAAGTCAA	26
I9F(-118)M13	Intron 9	TGTAAAACGACGGCCAGTCAGTGTGTAAGGATTAGTGCTGGT	27
I10R79M13	Intron 10	CAGGAAACAGCTATGACCGTCTGGGTGACAGAGAGAGACTCCA	28
I10F(-349)M13	Intron 10	TGTAAAACGACGGCCAGTGAGAAGTCGTTTTAGCATTTCCGTAT	29
DR1228M13	3'UTR	CAGGAAACAGCTATGACCCATTTGTTGTTTCTTATGGTCTGTGCTAT	30

*underlined nucleotides indicate M13 tag

Recombinant ASMT Expression Constructs and Allozyme Expression: *ASMT* cDNA sequences for the three non-synonymous cSNPs observed during the resequencing experiments were created using the QuickChange Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA), with the wild type *ASMT* cDNA open reading frame (ORF) in the pUni/V5-His-TOPO (pUni) vector (Invitrogen) as template. Specifically, the full-length wild type ORF was amplified using human brain Marathon-Ready cDNA (Clontech) as template. The resultant *ASMT* cDNA was subcloned into pUni, a vector that is only 2.3 kilobases in length and thus is well suited for performing "circular PCR" during site-directed mutagenesis. Site-directed mutagenesis was performed using internal primers that contained the variant nucleotide sequences. The *ASMT* cDNA inserts in pUni were excised and re-ligated into the eukaryotic expression vector p91023(b) (Wong et al. Science 228:810-815 (1985)). The sequences of inserts in p91023(b) were confirmed by completely sequencing both strands.

Expression constructs for the wild type and variant ASMT sequences were transfected into COS-1 cells using the TransFast™ reagent (Promega), with a 1:1 charge ratio. pSV-β-Galactosidase (Promega) was co-transfected as an internal control to make it possible to correct for transfection efficiency. The COS-1 cells were harvested after 48 hours and homogenized with a Polytron homogenizer (Brinkmann Instruments, Westbury, NY) in 25 mM potassium phosphate buffer, pH 7.8 containing 1 mM dithiothreitol (DTT) and 1 mM EDTA. Cell homogenates were centrifuged at 15,000 x g for 15 minutes, and the resultant supernatant preparations were used for enzyme assays and substrate kinetic studies.

ASMT *Enzyme Activity*: ASMT activity was measured using the methods of Zakharayan et al. (Chem. Res. Toxicol. 8(8):1029-1038 (1995)). Briefly, 0.10 M Tris-HCl buffer (pH 8.0), 4 mM glutathione, 1 mM magnesium chloride, 2.5 mM sodium arsenite, 10μM [^{14}C-CH_3]-S-adenosylmethionine, and recombinant enzyme were combined in a final volume of 250 μL. The cell homogenate preparations of recombinant ASMT allozymes described above were used for the activity studies without any further purification. The protein concentration of each recombinant protein preparation was determined by the dye-binding method of Bradford (Anal. Biochem. 72:248-254 (1976)) with bovine serum albumin as a standard. "Blank" samples included the same quantity of

COS-1 15,000 x g supernatant from cells that were transfected with "empty" p91023(b) expression vector to correct for endogenous activity.

Reaction mixtures were incubated for 60 minutes at 37°C and stopped by the addition of 750 μL 12 M HCl. The methylated arsenic compounds, products of the enzyme reaction, were isolated using the standard extraction procedure from Zakharyan et al., 1995, *supra*. Activities of recombinant ASMT allozymes were compared after correction for transfection efficiency by measuring the activity of cotransfected β-galactosidase using the β-Galactosidase Assay System (Promega) as described by the manufacturer.

Estimating Apparent K_m Values: To estimate apparent K_m values of ASMT for the sodium arsenite and SAM, a series of sodium arsenite and SAM concentrations were tested with the recombinant allozymes. Blanks for each substrate concentration were included by assaying COS-1 cell cytosol after transfection with empty p91023(b) vector. These data were fitted to a series of kinetic models, and the most appropriate model was selected on the basis of the dispersion of residuals and a determination of whether the *F*-test showed a significant reduction ($P < 0.05$) in the residual sums of squares. Apparent K_m values were calculated using the method of Wilkinson with a computer program written by Cleland. Wilkinson *supra*; and Cleland *supra*.

Western Blot Analysis: Quantitative Western blot analysis was performed with recombinant ASMT allozymes after expression in COS-1 cells. Polyclonal antibodies were generated against two synthetic polypeptides corresponding to ASMT amino acid residues 5-28 (RDAEIQKDVQTYYGQVLKRSADLQC; SEQ ID NO:31) and amino acid residues 341-360 (DIITDPFKLAEESDSMKSRC; SEQ ID NO:32). These antibodies were used to measure levels of immunoreactive ASMT protein with the ECL detection system (Amersham Pharmacia, Piscataway, NJ). The quantity of COS-1 cell preparation loaded on the gel for each allozyme was adjusted to achieve equal quantities of β-galactosidase activity, i.e., gel loading was adjusted to correct for transfection efficiency. The AMBIS Radioanalytic Imaging System, Quant Probe Version 4.31 (Ambis, Inc., San Diego, CA) was used to quantitate immunoreactive protein in each lane, and those data were expressed as a percentage of the intensity of the wild type ASMT band on the gel.

Reporter Activity: VNTR sequences from the *ASMT* 5'-UTR were subcloned into the pGL3-basic luciferase reporter vector (Promega). The VNTR sequences were prepared by PCR amplification, using individual DNA samples from subjects with known VNTR genotypes as templates. PCR primers had the sequences 5'-AAGAAGGGTACCACGAGATTTATCCGTGAAAATCGCA-3' (SEQ ID NO:33) and AAGAAGCTCGAGAGGGAAGGGGCTGGGGGCT (SEQ ID NO:34). PCR products and pGL3-basic were digested with *Xho*I and *Acc*65I and ligated together.

Reporters were cotransfected into HepG2 and HEK293 cells (American Type Culture Collection, Manassas, VA) using the TransFast™ reagent (Promega). The pRL-TK vector (Promega) was cotransfected as an internal control to correct for transfection efficiency. After 48 hours, cells were lysed and luciferase activity was measured using the Dual-Luciferase® Reporter Assay System (Promega).

Data Analysis: Statistical comparisons of data was performed by ANOVA with the StatView program, version 4.5 (Abacus Concepts, Inc., Berkeley, CA). Linkage analysis was performed after all DNA samples had been genotyped at each of the polymorphic sites observed, using the EH program developed by Terwilliger and Ott, Handbook of Human Genetic Linkage, The Johns Hopkins University Press, Baltimore, pp. 188-193 (1994). D' values, a quantitative method for reporting linkage data that is independent of allele frequency (Hartl and Clark Principles of Population Genetics, 3rd edition, Sinauer Associates, Inc., (Sunderland, MA), pp 96-106 (1997); and Hedrick Genetics of Populations, 2nd edition, Jones and Bartlett (Sudbury, MA), pp. 396-405 (2000)), were then calculated. The genotype data also were used to assign inferred haplotypes using a program based on the E-M algorithm (Long et al. Am. J. Hum. Genet. 56:799-810 (1995); and Excoffier and Slatkin Mol. Biol. Evol. 12:921-927 (1995)). Unambiguous haplotype assignment also was possible on the basis of genotype for samples that contained no more than one heterozygous polymorphism.

Example 2 – ASMT Polymorphisms

Eleven separate PCR amplifications were performed for each of the 120 DNA samples studied. All PCR amplicons were sequenced on both strands, making it possible to verify the presence of polymorphisms using data from the complimentary strand. A

total of 26 polymorphisms were observed (Table 2). Polymorphisms in exons, untranslated regions (UTR), and flanking regions (FR) are numbered relative to the adenine in the *ASMT* translation initiation codon (ATG, adenine is +1). Polymorphisms in introns are numbered separately, either as positive numbers relative to the guanine in the splice donor site (GT, guanine is +1), or as negative numbers relative to the guanine in the splice acceptor site (AG, guanine is -1).

Variant allele frequencies ranged from 0.8% to 45%, with differences between the African-American and Caucasian-American subjects. Twenty-two polymorphisms were observed in 60 DNA samples from African-American subjects, while 21 were found in the 60 samples from Caucasian-American subjects. The overall number of *ASMT* polymorphisms per kilobase of sequence in the 120 samples studied (4.8 polymorphisms/kilobase, Table 3) was close to that (4.6/kilobase) observed in similar studies of other human genes (Halushka et al., Nature Genet., 22:239-247 (1999)). Three of the SNPs were within the coding-region (cSNPs). All of these were nonsynonymous and resulted in the amino acid alterations Arg173Trp, Met287Thr, and Thr306Ile. The Arg173Trp polymorphism had a frequency of 0.8% in African-American subjects but was not observed in DNA from Caucasian subjects. The Met287Thr polymorphism had a frequency of 10.8% in African Americans, and 10% in Caucasians. The Thr306Ile polymorphism had a frequency of 0.8% in Caucasians but was not observed in DNA from African-American subjects. To exclude artifacts introduced by PCR-dependent misincorporation, independent amplifications were performed and the amplicons were sequenced in all cases in which a polymorphism was observed only once among the DNA samples studied.

TABLE 2
ASMT Polymorphisms

Location	Nucleotide	Wild Type	Variant	Amino Acid	African American	Caucasian American
5'-FR	-676	T	C		0.000	0.017
5'-FR	-542	G	A		0.025	0.000
5'-FR	-477	A	G		0.317	0.383
5'-FR	-420	C	G		0.008	0.000
5'-FR	-339	T	C		0.108	0.217
5'-FR	-116	C	A		0.008	0.000
5'-FR	-114	G	C		0.242	0.117
Intron 2	I2(-75)	T	A		0.025	0.008
Intron 2	I2(-47)	D	I	C insertion	0.267	0.208
Intron 2	I2(-10)	G	T		0.108	0.017
Intron 3	I3(-18)	G	A		0.000	0.008
Intron 4	I4(217)	A	G		0.058	0.008
Intron 4	I4(365)	G	A		0.042	0.000
Intron 4	I4(414)	T	C		0.100	0.008
Intron 4	I4(467)	A	T		0.067	0.008
Exon 6	517	C	T	Arg173Trp	0.008	0.000
Intron 6	I6(56)	A	G		0.108	0.100
Intron 6	I6(-56)	G	C		0.217	0.242
Intron 8	I8(154)	A	C		0.108	0.117
Intron 8	I8(213)	C	T		0.092	0.142
Exon 9	860	T	C	Met287Thr	0.108	0.100
Intron 9	I9(-50)	T	C		0.183	0.217
Exon 10	917	C	T	Thr306Ile	0.000	0.008
Intron 10	I10(-282)	C	T		0.267	0.350
Intron 10	I10(-189)	G	A		0.000	0.025
Intron 10	I10(-94)	G	A		0.375	0.450

TABLE 3
ASMT polymorphism frequencies

	Total	African-American	Caucasian-American
SNPs/Kb	4.8	4.0	3.8
SNPs/Kb coding	2.7	1.8	1.8
SNPs/Kb non-coding	5.3	4.6	4.4
SNPs/Kb UTR	41.2	35.3	23.5
SNPs/Kb Intron	4.7	4.1	4.4
nonsyn/kb	0.5	0.4	0.4

Example 3 - Linkage disequilibrium and haplotype analysis

Linkage disequilibrium analysis was performed after all of the DNA samples had been genotyped at each of the polymorphic sites. Pairwise combinations of these polymorphisms were tested for linkage disequilibrium using the EH program developed by Terwilliger and Ott, Handbook of Human Genetic Linkage, The Johns Hopkins University Press, Baltimore, pp. 188-193 (1994). The output of this program was used to calculate D' values, a method for reporting linkage data that is independent of sample size. All pairwise combinations with a linkage disequilibrium greater than or equal to 1 in at least one population are shown in Table 4.

Twenty-two unequivocal haplotypes were identified by these studies (Table 5). The unequivocal haplotypes included seven haplotypes that were common to both ethnic groups, and fifteen that were ethnic specific (seven haplotypes were specific for African-American subjects; eight haplotypes were specific for Caucasian subjects).

Table 4
ASMT linkage disequilibrium analysis

AA Polymorphism pair		D' Value	p value
-477	-339	1	0.000019
-477	-114	0.7652	0
-477	I6(56)	1	0.000019
-477	860	1	0.000019
-477	I10(-94)	1	0
-339	-114	1	0
-339	I6(56)	1	0
-339	860	1	0
-339	I10(94)	1	0.000151
-114	I2(-10)	1	0
-114	I4(217)	1	0.000072
-114	I4(414)	1	0
-114	I4(467)	1	0.000026
-114	I6(56)	1	0
-114	860	1	0
-114	I10(-94)	1	0
I2(-75)	I8(213)	1	0.000907
I2(-47)	I8(213)	1	0.000036

AA Polymorphism pair		D' Value	p value
I2(-10)	I4(217)	1	0
I2(-10)	I4(365)	1	0.000007
I2(-10)	I4(414)	1	0
I2(-10)	I4(467)	1	0
I2(-10)	I10(-282)	1	0.000001
I2(-10)	I10(-94)	1	0.000073
I4(217)	I4(365)	1	0
I4(217)	I4(414)	1	0
I4(217)	I4(467)	1	0
I4(217)	I6(-56)	1	0.000028
I4(217)	I9(-50)	1	0.000007
I4(217)	I10(-282)	1	0.000164
I4(365)	I4(414)	1	0.000006
I4(365)	I4(467)	1	0
I4(365)	I6(-56)	1	0.000545
I4(365)	I9(-50)	1	0.0002
I4(414)	I4(467)	1	0
I4(414)	I6(-56)	1	0
I4(414)	I10(-282)	1	0.000001
I4(414)	I10(-94)	1	0.000124
I4(467)	I6(-56)	1	0.000009
I4(467)	I9(-50)	1	0.000002
I4(467)	I10(-282)	1	0.000068
I4(467)	I10(-94)	1	0.000855
I6(56)	860	1	0
I6(56)	I10(-94)	1	0.000151
I6(-56)	I8(213)	1	0.000151
I6(-56)	I10(-282)	1	0
I6(-56)	I10(-94)	1	0
I8(213)	I10(-282)	1	0.000008
I8(213)	I10(-94)	1	0.000911
860	I10(-94)	1	0.000151
I9(-50)	I10(-94)	1	0
I10(-282)	I10(-94)	1	0

CA Polymorphism pair		D' Value	p value
-477	-339	0.9276	0
-477	-114	1	0.000014
-477	I6(56)	1	0.000082
-477	860	1	0.000082
-477	I10(-94)	0.9155	0
-339	-114	0.7938	0.000012
-339	I6(56)	0.9773	0.000011
-339	860	0.8773	0.000011
-339	I9(-50)	0.4721	0.000063
-339	I10(94)	0.8307	0.000003
-114	I6(56)	1	0
-114	860	1	0
I6(56)	860	0.9065	0
I6(-56)	I8(213)	1	0
I6(-56)	I10(-282)	0.9385	0
I6(-56)	I10(-94)	0.9185	0
I8(213)	I10(-282)	1	0
I8(213)	I10(-94)	1	0.000042
860	I10(-94)	1	0.000195
I9(-50)	I10(-94)	0.913	0
I10(-282)	I10(-94)	1	0

Table 5

ASMT haplotype analysis

AA Freq.	CA Freq.	-676	-477	-339	-114	I2(-75)	I2(-47)	I2(-10)	I3(-18)	I4(217)	I4(365)	I4(414)	I4(467)	I6(56)	I6(-56)	I8(154)	I8(213)	860	I9(-50)	I10(-282)	I10(-94)
0.5530	0.4365	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
0.0661	0.0276	WT	V	V	V	WT	V	WT	WT	WT	WT	WT	WT	V	WT	WT	WT	V	WT	WT	V
0.0652	0.0599	WT	V	WT	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	V	WT	V	WT	WT	V	V
0.0545	0.0622	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	WT	WT	WT	WT	WT
0.0339	0.0264	WT	V	V	V	WT	WT	WT	WT	WT	WT	WT	WT	V	WT	WT	WT	V	WT	WT	V
0.0277	0.0000	WT	V	WT	V	WT	V	V	WT	WT	WT	V	WT	WT	V	WT	WT	WT	V	V	V
0.0158	0.0000	WT	V	WT	V	WT	V	V	WT	V	V	V	V	WT	V	V	WT	WT	V	V	V
0.0100	0.1667	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	V	V
0.0098	0.0000	WT	V	WT	WT	V	V	WT	WT	WT	WT	WT	WT	WT	V	WT	V	WT	WT	V	V
0.0092	0.0000	WT	WT	WT	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
0.0091	0.0000	WT	V	WT	V	WT	V	V	WT	V	V	V	V	WT	V	WT	WT	WT	V	V	V
0.0091	0.0000	WT	V	WT	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	V	V
0.0085	0.0833	WT	V	WT	V	WT	WT	V	WT	V	V	V	V	WT	V	WT	WT	WT	V	V	V
0.0083	0.0000	WT	WT	WT	V	WT	WT	V	WT	V	V	V	V	WT	V	WT	WT	WT	V	V	V
0.0000	0.0667	WT	V	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	V	V
0.0000	0.0583	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	WT	V	WT	WT	V	V
0.0000	0.0470	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	WT	WT	WT	V	V	V
0.0000	0.0127	WT	V	V	V	WT	V	WT	WT	WT	WT	WT	WT	V	WT	V	WT	V	WT	WT	V
0.0000	0.0097	WT	V	V	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	V	V
0.0000	0.0097	WT	WT	WT	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	V	WT	WT	WT	V	V	V
0.0000	0.0097	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
0.0000	0.0083	WT	V	V	WT	WT	WT	WT	V	WT	WT	WT	WT	WT	WT	WT	WT	WT	V	V	V

Example 4 – Activity of ASMT allozymes

Cell homogenate preparations containing recombinant ASMT allozymes, prepared from six independent COS-1 cell transfections as described in Example 1, were used to assess catalytic activity. The resulting activity was adjusted to a percentage of the WT ASMT enzyme activity, shown as mean ± SEM in Table 6. The activities of Arg173Trp, Met287Thr, and Thr306Ile were 31%, 350%, and 3.2% that of the WT ASMT enzyme, respectively. Western blotting revealed that the protein levels of the three allozymes were 20%, 190%, and 1.1% that of the WT ASMT enzyme, respectively. Thus, the effect of each cSNP on enzyme activity was at least partially accounted for by the effect on the protein level.

Alterations in amino acid sequence can alter enzyme substrate affinity and/or catalytic efficiency. Substrate kinetic studies were conducted to determine whether the Arg173Trp, Met287Thr, and Thr306Ile allozymes differed from the WT ASMT protein in these aspects. A series of sodium arsenite and SAM concentrations were used to estimate apparent K_m values for recombinant wild type ASMT and for the three variant allozymes. These studies revealed a significant difference in apparent K_m values between the WT ASMT protein and the Met287Thr allozyme for sodium arsenite (4.6 μM vs. 11 μM, respectively, $P < 0.05$; Table 6). There was no significant difference in apparent K_m values between the WT protein and the Arg173Trp allozyme, and the Thr306Il3 allozyme was not used in kinetic studies.

Table 6
Human ASMT allozyme activity

Allozyme	Enzyme Activity (% WT)	Immunoreactive protein (% WT)	Sodium Arsenite Km, μM	SAM Km, μM
WT	100	100	4.6 ± 0.56	12 ± 6.9
Arg173Trp	31 ± 2.6**	20 ± 0.5**	3.1 ± 0.8	8.9 ± 1.2
Met287Thr	350 ± 89*	190 ± 14*	11 ± 1.8*	4.5 ± 0.9
Thr306Ile	3.2 ± 2.1**	1.1 ± 0.6**	ND	ND

*, $P < 0.05$; **, $P < 0.001$; ND, not determined

Example 5 – *ASMT* 5'-UTR VNTR reporter activity

Further examination of the *ASMT* 5' flanking region revealed the presence of two or more VNTRs in the DNA from each subject. As shown in Table 7, Caucasian American subjects contained two or three VNTRs, while African American subjects contained two, three, or four VNTRs. The majority of subjects contained one two-repeat allele and one three-repeat allele, or were homozygous for the three-repeat allele.

Table 7
Human ASMT 5'-UTR VNTR

	Allele Frequencies	
Repeat Number	**CA**	**AA**
2 (**V2*)	0.375	0.375
3 (**V3*)	0.625	0.558
4 (**V4*)	0	0.067
	Genotype Frequencies	
Genotype	**CA**	**AA**
**V2/*V2*	0.183	0.100
**V2/*V3*	0.383	0.433
**V2/*V4*	0	0.117
**V3/*V3*	0.433	0.333
**V3/*V4*	0	0.017
**V4/*V4*	0	0

Each VNTR had one of two similar but not completely identical nucleotide sequences. The first sequence, designated as subunit A, was 36 nucleotides in length and had the sequence 5'-GTCGCAGGCCGAGGAGACAGTGAGTGCGCGCCCTGA-3' (SEQ ID NO:35) from the 5'-FR and exon 1. The second sequence, designated as subunit B, was 35 nucleotides in length and had the sequence 5'-GTCGCAGGCCGAGGAGACAGTGAGTGCGCGCCCTG-3' (SEQ ID NO:36) from exon 1 and intron 1. The DNA from each subject included at least one VNTR having the sequence of subunit B.

Reporter constructs containing two, three, or four *ASMT* VNTRs were transfected into HepG2 and HEK293 cells as described in Example 1. Luciferase activity was measured after 48 hours of culture. The levels of reporter activity were expressed as the percent activity of the basic (i.e., empty) vector control, and are shown in Figures 4A and

4B. Construct names refer to repeat numbers, with *V2 containing one copy of subunit A and one copy of subunit B (configured as 5'-AB-3'), *V3 containing two copies of subunit A and one copy of subunit B (configured as 5'-AAB-3'), and *V4 containing three copies of subunit A and one copy of subunit B (configured as 5'-AAAB-3'). Construct *2V contained two repeats (configured as 5'-AB-3') and had a substitution of cytosine for guanine at nucleotide -114. The variant at -114 was only observed in alleles with two VNTR subunits. The values graphed in Figs. 4A and 4B represent mean ± SEM (n = 4).

In these studies, all of the reporters resulted in luciferase activity that was dramatically increased as compared to that of the vector control. In HepG2 cells (Fig. 4A), a decrease in the number of VNTRs resulted in gradually increased reporter activity. The construct containing four VNTRs had significantly less activity than those containing three or two VNTRs. The combination of two VNTRs with a cytosine at position -114 (reporter *2V) resulted in the greatest level of luciferase activity. The opposite effect was observed in HEK293 cells, where the reporter activity of the construct containing four VNTRs was significantly greater than the activity of the other constructs. Thus, the effect of repeat number may be determined by factors specific to each type of cell.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.